Quantification of uncertainties in fossil leaf aleoaltimetry: does leaf size matter?

Quantification of Uncertainties in Fossil Leaf Paleoaltimetry - Does Leaf Size Matter?

Robert A. Spicer [1, 2] and Jian Yang [2]

[1] Department of Earth and Environmental Sciences, Centre for Earth, Planetary, Space and Astronomical Research, The Open University, Milton Keynes, MK7 6AA, UK

[2] Institute of Botany Chinese Academy of Sciences,
No.20 Nanxincun, Xiangshan, Beijing 100093, China.

Abstract

The utility of multivariate foliar physiognomy, specifically the Climate Leaf Analysis Multivariate Program (CLAMP), to yield reliable estimates of enthalpy and hence paleoelevation has been demonstrated by comparison with other proxies, yet concerns have arisen regarding uncertainties arising from 1) apparent ambiguities in the scoring regime and 2) the way leaf size is scored. Regarding 1) scoring ambiguities are examined by reporting on scoring tests with novice users and inter-laboratory comparisons. The uncertainties were found to be less than those arising from the statistical methodology underpinning CLAMP. In respect of 2) the effect of removing all size data both from modern test sites and fossil data was tested. Specifically the effect of removing leaf size data from the 15 Ma Namling data set from south central Tibet, was investigated. Removal of all size data from modern sites demonstrated that size data contributes little to estimates of MAT (Mean Annual Temperature) and enthalpy. Similarly the removal of leaf size information from the Namling data set alone, but with calibration unchanged, and from both the Namling site and

calibration sites, this time with recalibration, still yield paleoelevation estimates that have been independently matched by oxygen isotope techniques. Moreover the removal of all leaf size information results in only small increases in uncertainty (± 52 m).

1. Introduction

Paleoaltimetry plays an important role in understanding crustal and mantle dynamics and several methods of measuring surface height (as distinct from uplift of rocks or exhumation [England and Molnar, 1990]) have emerged and include: 1) methods based on the architecture of woody dicotyledonous leaves (foliar physiognomy) [e.g. Wolfe et al. 1997, 1998; Spicer et al., 2003], 2) oxygen isotopes, [e.g. Garzione et al. 2000; Rowley et al. 2001; Rowley and Garzione, 2007; Quade et al. 2007], 3) ^{13}C - ^{18}O bonds [Gosh et al., 2006], 4) *D/H* ratios of plant lipid biomarkers [Polissar et al. 2009], pCO_2 [McElwain, 2004] and 5) atmospheric pressure [Sahagian et al., 2002]. Where multiple methods have been applied in the same basin and to rocks of the same age, CLAMP (Climate Leaf Multivariate Analysis Program) and oxygen isotopes yield very similar height estimates [Spicer et al., 2003; Currie et al., 2005; Rowley and Currie, 2006], as do lipid biomarkers and oxygen isotopes [Polissar et al., 2009]. Clearly precision is crucial here and minimizing uncertainties in height estimates (improving precision), as well as accuracy, is a goal of all methods. Because all methods carry caveats a multiproxy approach is essential in any paleoelevation study.

Peppe et al. [2010] question the precision of CLAMP based on their contention that leaf size is inappropriately scored in the methodology. They attempt to offer a different methodology and their results, at first sight, appear rather damning regarding the way that leaf size has been incorporated into the CLAMP calibration. However we question their approach. Our concerns regarding the validity of their assertions fall into two categories: 1) failure to adhere to published CLAMP protocols and 2) failure to recalibrate CLAMP when scoring test sites differently from the

calibration sites. Published data are insufficient to replicate or correct their results, but it is possible to test their assertion by simply removing all leaf size data from the analysis and examining the changes in precision and accuracy in paleoelevation estimates that result.

1.1. Foliar Physiognomy and Paleoelevation

In respect of foliar physiognomy a univariate approach for determining mean annual temperature (MAT) has a long history and exploits the observed correlation between the proportion of toothed versus un-toothed leaf margins to temperature [Bailey and Sinnott, 1915, 1916; Wolfe, 1979; Wilf, 1997]. In the context of paleoaltimetry, temperature (usually MAT) and assumptions of how temperature varies with surface height across different regions (often erroneously referred to as "lapse rates"; a term that should be restricted to free air changes in temperature with altitude) have previously been pursued to obtain a direct estimate of surface height [e.g. Axelrod, 1968; Meyer, 1992]. Unfortunately surface temperatures do not depend simply on elevation, but also on surrounding elevations, patterns of atmospheric circulation, specific humidity, and other parameters that are not easily predicted for past times so this approach can lack precision [Wolfe, 1992].

With the advent of increased computing power a more sophisticated multivariate approach was developed known as CLAMP [Wolfe, 1993] whereby enthalpy can be derived from foliar physiognomy. Enthalpy (H) is defined as follows:

$$(1)\quad H = c_p T + L_v q$$

where c_p is the specific heat capacity at constant pressure of moist air (J/kg/K), T is temperature (K), L_v is the latent heat of vaporization of water (J/kg) and q is the specific humidity (g/kg). Enthalpy is strongly coded in leaf form because of its relationship with moist static energy, which is a function of both temperature and moisture; two environmental variables important to plant growth.

Moist static energy (h) is the sum of enthalpy and the gravitational potential energy (gZ) where g is the acceleration due to gravity and Z is the elevation. Thus the difference in enthalpy between a site at known elevation and another site where elevation is unknown, divided by g, will yield the height difference between the two sites [Wolfe et al., 1997, 1998; Forest et al., 1999; Spicer et al., 2003].

CLAMP capitalizes on the observation that leaf architecture is invariably a compromise solution between the conflicting constraints, in any given environmental setting, of maximizing photosynthetic efficiency while simultaneously minimizing structural investment and other constraints such as water loss. Physiognomic environmental adaptation occurs within the capabilities imparted by the genome honed by long-term natural selection. Non-adapted physiognomies fail to survive, and over time successful physiognomies emerge largely independent of taxonomy [Spicer, 2000, 2007, 2008]. This is easily demonstrated by the convergence of leaf form in spatially and taxonomically distinct regions, but which experience similar climatic regimes. Examples include the similar foliar physiognomies seen in deserts in Asia and the Americas, or indeed convergent leaf form between the rain forests in those areas. Similarly the vegetation of Chile, California and Southern Europe, all of which experience a Mediterranean type climate, display closely allied physiognomies. Undoubtedly there is a biogeographic component to physiognomy as demonstrated in the context of CLAMP by Kennedy at al. [2002] and from more general observations [Jordan, 1997; Doyle, 2007; Little et al., 2008], but this does not negate the derivation of a climate signal that clearly exists in leaf form. If this were the case then the long established univariate approach would fail to show a correlation with climate.

1.2. Why a multivariate approach?

Adaptive success is not determined by any single architectural feature [Lande and Arnold, 1983], and no single feature can be expected to correlate with a single climatic variable. Fitness is influenced instead by numerous interacting traits [Ackerly et al., 2000] spanning the architectures of conducting tissues in roots, stems and leaves, overall canopy architecture [Hellicker and Richter, 2008] and foliar physiognomy, but the nature of plant organ shedding and fossilization means that isolated leaves provide the only commonly available measure of this plant/climate relationship in the fossil record.

A useful analogy when considering the multivariate nature of leaf form is that of a racing car in which the engine, brakes, tires, body shape and driver skill all contribute to competitive success. No single attribute operating alone can deliver that success. Some attributes such as engine and braking power are even antithetic. So it is with leaves: the need to conserve water and minimize structural investment, and thus have small leaves, is antithetic to having large surface areas with which to intercept light. The result is, through selection, an engineering solution that is a compromise. With this in mind the inappropriateness of seeking a correlation between a single leaf character and a single environmental variable is self-evident. The multivariate nature of the problem also means that seeking a simple ecophysiological explanation for any given character is largely futile and this may well be the explanation why simple relationships linking single leaf characters with single environmental variables have not been found. Far from negating the CLAMP methodology as Peppe et al. [2010] imply, this re-enforces the need for a multivariate approach. Moreover, in so far as temperature, radiation balance, diffusion, evaporation and precipitation are all correlated in the climate system, techniques for analyzing the relationship between leaf architecture to climate not only have to be multivariate, but have to accommodate attributes and variables that are also strongly correlated.

Enthalpy is a good example of where changes in both temperature and humidity are strongly correlated and in turn are correlated with altitude. To attempt to deconstruct the multivariate nature of leaf physiognomy in relation to enthalpy by looking at the effect of a single leaf attribute such as size is not useful.

That said there are several procedural problems with the Peppe et al. [2010] approach that can be categorized as 1) not following the CLAMP protocols and 2) not understanding the nature of calibration. Before examining these procedural issues more closely it is useful to understand the principles of the CLAMP approach.

1.3. CLAMP Methodology

CLAMP methodology has evolved since the introductory paper of Wolfe [1993]. The calibration data set offering the most precision and most applicable to warmer climates of the past is that now known as PHYSG3BRC, which consists of 144 vegetation samples scored for 31 physiognomic character states grouped into 7 categories (lobing, margin type, size, apex and base morphologies, length to width ratios and shape). Wolfe's original dataset consisted of 106 samples (later reduced to 103) scored for 29 character states. The two additional states were the inclusion of a new size category for the smallest leaves (Nanophyll) and another for the largest (Mesophyll 3). Accompanying the PHYSG3BRC physiognomic data set is a corresponding meteorological data file for the same 144 sites, but containing a minimum of 30 year normals for Mean Annual Temperature (MAT), Warm Month and Cold Month Mean Temperatures (WMMT and CMMT), Length of the Growing Season (LGS), Growing Season Precipitation (GSP) Mean Monthly Growing Season Precipitation (MMGSP), precipitation on the three wettest and three driest months (3WET and 3DRY), Relative Humidity (RH), Specific Humidity (SH) and Enthalpy. Both PHYSG3BRC and

MET3BR are available form the CLAMP website (http://www.open.ac.uk/earth-research/spicer/CLAMP/Clampset1.html).

Canonical Correspondence Analysis (CCA) [Ter Braak, 1986] is used to arrange the vegetation sites in multidimensional space using the physiognomic data, and the climate data are used to align and calibrate climate vectors running through this space. CCA is used because it is robust to missing data and makes no assumptions about the independence of the variables. It is quite clear that leaf characters are not independent of one another just as climate variables are correlated through the physics of the atmosphere. That said, CCA is not perfect as there is a theoretical requirement for unimodality in the distribution of leaf characters with respect to environmental gradients, but even here it seems remarkably robust in practice and only begins to produce unreliable results when numerous characters exhibit more complex response patterns such as bimodality [ter Braak, 1986].

1.4. CLAMP Protocols

In order to construct an internally consistent calibration (training) data set Wolfe [1993] established the following protocols:

1) CLAMP calibration samples should reflect the overall morphological variation seen in the leaves of each species of woody dicot present in the calibration site. This morphological variation includes that displayed across all 31 leaf character states including size categories.
2) All woody dicots should be samples from all growth habits: trees, bushes and vines. In the fossil record it is often impossible to determine the habit of the source plant so all habits are included in the calibration.

3) The scoring protocols were empirically devised by Wolfe [1993] and designed to reflect the kind of leaf material most often found in the fossil record. It was for this reason that an upper limit to the size categories used in calibration was chosen. Large leaves rarely survive intact during the pre-depositional transport, taphonomic and subsequent collecting process. That large leaves occasionally are preserved and recovered is irrelevant as far as the calibration process is concerned.
4) A minimum of 20 woody dicot species is recommended for both the calibration and fossil sites. This minimum diversity requirement was tested empirically [Povey et al. 1994] using an unusually rich flora (>130 species) from the Puget Group, Washington, U.S.A.. With less than 20 species uncertainties rose markedly, while diversities greater than 30 species yielded negligible increases in precision.

These protocols are clearly stated on the CLAMP website and yet several test procedures used by Peppe et al. [2010] do not meet these criteria. Some of their test sites from Central Connecticut had diversities as low as 13 and all were below the 20 species minimum. The inclusion of such low diversity sites has the effect of inflating the uncertainties [Povey et al.1994]. Insufficient detail is supplied with other sites to assess whether or not sampling protocols were followed. Within each sample just a few leaves (sometimes only 2) were selected randomly for measurements that were supposed to represent the full range of morphological variation present in the sample. A better strategy would have been to measure all the available leaves in both test and vouchered calibration samples so that new calibrations could have been constructed.

Underlying the CLAMP protocols is an appreciation of how potential fossil leaf assemblages are filtered by taphonomic processes. Thus Wolfe deliberately devised a calibration scheme that did not assume the exceptional preservation of large leaves. Such leaves experience higher mechanical stresses per unit area than smaller leaves and are preferentially mechanically degraded during pre-

depositional transport processes. The inclusion of large leaves in a calibration, when only the smaller leaves within the source population might be expected to be preserved as fossils, would prejudice the analysis of the paleoenvironment against any variable that was dependent on large leaf sizes. Thus the statement by Peppe et al. [2010] that "fossil sites measured following CLAMP methodology [e.g. Wolfe et al., 1997; Kennedy et al., 2002; Spicer et al., 2003] would [also] be variably biased to underestimate leaf area, and thus also incorrectly estimate climate variables" betrays a misunderstanding of the nature of calibration. The statement would only be correct if fossil leaf assemblages routinely preserved the largest leaves in the source vegetation, which they undoubtedly do not. The open-ended 'Mesophyll 3' size category allows for large leaves, but does not depend on their presence for reliable paleoclimate estimates.

1.5 CLAMP and the need for Unimodal Distribution of Leaf Characters

Peppe et al. [2010] are criticize CLAMP in part because the theoretical model underpinning CCA assumes a unimodal response of leaf characters to environmental constraints. In plant ecology where CCA is the ordination method of choice, a unimodal Gaussian response of a species to an environmental gradient is assumed. The forerunner of CCA, Correspondence Analysis (Benzecri, 1973; Hill, 1973, 1974) has been shown by Ter Braak [1985] to approximate the maximum likelihood solution of Gaussian ordination if the sampling distribution of the species abundances is Poisson, and if the following conditions (C1-C4) are met:

C1 - the species' tolerances (responses in relation to an environmental gradient) are equal,

C2 - the species' maxima (maximum responses, for example abundance, in relation to an environmental gradient) are equal,

C3 - the species' optima (positions along the gradient where the greatest responses occur) are homogeneously distributed over an interval A (range of the environmental gradient) that is large compared to tolerance,

C4 - the site scores are homogeneously distributed over a large interval B that is contained in A.

According to Ter Braak [1986, p. 1168] “conditions C1 and C2 are not likely to hold in most natural communities, but the usefulness of correspondence analysis in practice relies on its robustness against violations of these conditions [Hill and Gauch, 1980]” and concludes that CCA “derives its theoretical strength from its relation to maximum likelihood Gaussian canonical ordination under conditions C1-C4 and furthermore seems extremely robust in practice when these assumptions do not hold” [Ter Braak, 1986, p. 1177]. In respect of unimodality Ter Braak [1986. p. 1177] stated “the model would not work if a large number of species were distributed in a more complex way, e.g. bimodally; the restriction to a unimodal model is necessary for practical solubility”. Here the emphasis is on the extent to which bimodality is present in the data set. This assertion was tested by Palmer [1993] who demonstrated by the use of simulation studies that CCA is robust even in the face of skew and noise, while Hill [1991] used CCA to predict spatial distributions using binary data. The CLAMP categorical unitary scoring process tends to force compliance with conditions C1-C4 and especially C1 and C2, even when converted to the percentage scores for each site, but it does not eliminate bimodality.

From the above account of the use of CA and CCA in ecology it should be apparent that the terms “unimodal” and “bimodal” refer to the distribution of a species’ abundance along an environmental gradient as represented in the different sites. In CLAMP it is the distribution of leaf character states that equate to species distribution in ecological studies, but again the distribution has to be viewed in terms of possible environmental gradients as represented in the site score distribution along, for example, axis 1 of the ordination [Ter Braak, 1986] and not as given in Peppe et al. [2010] supplementary Figure 1, the distribution of site means. In CLAMP there is no evidence for a large number of leaf characters having bimodal distributions along any of the principle axes of variation or climate vectors. The observation that the regressions used to derive the paleoenvironmental data

are in the most part so “clean” (see the CLAMP website for details of the regressions) is further evidence that bimodality or other complex distributions is not compromising CLAMP CCA performance. However that is not to say that further improvements could not be made by transforming some of the scores.

1.6. Physiognomic Space and Calibration Issues.

Using the published CLAMP protocols, leaves from modern training sites are scored and CCA is used to position the training sites relative to one another in multidimensional space to create a cloud of reference points that has been termed “physiognomic space” [Spicer, 2007]. For any given training dataset this cloud has a well-defined geometry and boundaries. Fossil sites, or modern sites used for testing, that fall outside of this physiognomic space will inevitably be associated with higher uncertainties than sites that fall within the boundaries of calibrated physiognomic space. A primary step in any analysis is to see where an unknown resides with respect to these boundaries. An analogy here is a mercury thermometer. When the mercury meniscus is beyond the graduated scale there are likely to be increased and unquantified errors in determining temperature. Peppe et al. [2010] do not state if any of their tests sites meet this basic criterion.

The structure of physiognomic space is determined by the way that the leaves are scored. Any change in the scoring method will change the spatial relationship between the sites in multidimensional space. Climate vectors summarize the trends across physiognomic space in terms of the observed climate, measured either in or close to the vegetation stands, or by means of a globally gridded climate data set adjusted for local altitude [Spicer et al., 2009]. The vectors are calibrated using the observed climate relating to each vegetation site and plots are made of the vector score (position along the vector) against the observed climate. 2nd order polynomial regressions of the vector score against observed climate provide a model for estimating the climate

of a site where only the vector score is known (usually a fossil site). These regression plots are given on the CLAMP website. The scatter of the residuals about this regression provides a measure of the uncertainty of this estimate and reflects a composite uncertainty arising from the use of all the leaf characters. There is no attempt to relate any given character to any given climate variable. It follows from the correlated nature of multivariate constraints on leaf architecture that any such attempt to link specific leaf characters to specific climate variables would be futile and potentially misleading.

A consequence of this methodology is that any change in the scoring regime will result in a shift in the relative positions of the sites in physiognomic space, and thus the relationship to the vectors. If a recalibration is not carried out by constructing a new regression line, there will be an apparent, but erroneous, degradation in precision. Although not very clear, what appears to have happened in the Peppe et al. [2010] analysis is that the scoring regime in respect of size for test sites was altered by measuring size directly instead of following the scoring scheme. However this change in scoring was not applied to the calibration sites, and no new regression equations were calculated. Consequently the test sites were miss-positioned with respect to calibrated physiognomic space and the result was a set of spurious and invalid estimates of enthalpy. Peppe et al. [2010] claim that changes in the scoring protocols that result in changes to the structure of physiognomic space mean that CLAMP is “unstable”. To use the analogy of a mercury thermometer this would be the same as claiming that the thermometer is an unstable instrument because it gives a different reading when the graduated scale is moved, stretched or shortened, or the mercury was replaced by another liquid with a different expansion co-efficient.

2. Materials and Methods

2.1. Experimental Strategy

To evaluate Peppe et al's [2010] claims (but adhering to the CLAMP protocols and with appropriate re-calibration where necessary) a series of experiments were carried out using the most commonly used CLAMP calibration dataset PHYSG3BRC. This data set was used throughout for simplicity and because it was appropriate and common to all analyses.

1) The combined uncertainty arising from novice collecting and scoring was tested using eight pairs of students analyzing a modern test site near Beijing, China.
2) To see which leaf character suites had most influence in a multivariate context for determining MAT and Enthalpy accurately and precisely, 26 of the PHYSG3BRC calibration sites selected for their uniform distribution throughout physiognomic space were analyzed with various characters state suites removed.
3) Peppe et al's [2010] experiments using part of the PHYSG3BRC dataset were repeated, but not in their entirety. Some of their test sites belong to the PHYS3ARC dataset that include the so-called "alpine nest" and it is inappropriate to evaluate these using the PHYSG3BRC calibration. In this experiment all size data were removed from the test sites. This is equivalent to missing data in fossil sites and does not require recalibration.
4) The Namling data set of Spicer et al. [2003] from the south central Tibetan Plateau was re-analyzed without size information to estimate the difference from the original paleoelevation determinations that have been independently corroborated using oxygen isotopes [Currie et al., 2005]. Again this simulated catastrophic taphonomic loss of all size data from the fossil assemblage.
5) Experiment 4 was repeated but also with all size data removed from the PHYSG3BRC dataset as well as the Namling assemblage. This required recalibration.

It was not possible to replicate all of Peppe et al's [2010] experiments. Their Central Connecticut sites had to be rejected because they lack sufficient species to provide a fair test. The other sample sites from the Eastern US have been collected for other purposes and may not conform to CLAMP protocols. Finally Wolfe and Uemura [1999] cast doubt on the appropriateness of the Panama dataset, as well as many others, used in Wilf et al. [1998] and it remains unclear if this data set was assembled at least in part from herbarium material, or even floral lists, and therefore does not conform to CLAMP protocols.

2.2. Uncertainty arising from inaccurate scoring.

Early in CLAMP development it was recognized that although the technique does not require the identification of species (i.e. assigning them to a particular taxonomic group) it does depend on distinguishing one species from another even if these are given arbitrary names. This 'morphotyping' is usually carried out not only using leaf architecture but also on the basis of venation patterns. Only a minimal skill in botany is required for this as it is essentially a pattern recognition exercise, thus allowing the technique to be used widely.

More problematic is correct scoring of the species. Wilf [1997] was of the opinion that the character definitions were, as originally expressed, ambiguous in some cases. Subsequently this issue has been restated by other workers [Wiemann et al., 1998; Green, 2006]. As a result of these concerns the character definitions were re-written and the scoring procedure made more explicit on the CLAMP website. To test the revised scoring support documentation, and to quantify uncertainties arising from scoring by inexperienced users, eight pairs of novice student scorers were asked to make their own collections of leaves from the Xiaolongmen reserve near Beijing during a joint RCUK/CAS summer school held in October 2008. The total number of woody dicots in the woodland was above the minimum of 20 required and typically the students collected 23 species,

sampling their own perceived morphological range of leaf form within each species. The mixed UK and Chinese class also allowed us to evaluate trans-linguistic understanding of the scoring regime. All pairs were then led through the scoring of a single common taxon as a group exercise, but thereafter each pair was required to score their own collections, only seeking assistance when essential. Scoring was done using the then revised scoring documentation, but feedback from this exercise subsequently informed further revision. The student scoresheets were then analyzed using the standard CLAMP procedure and the outcome is presented in Table 1.

Recently Su Tao, Frédéric Jaques and others at the Institute of Botany Chinese Academy of Sciences in Kunming independently collected and scored 50 new CLAMP sites from modern vegetation across China (Jaques et al., pers. comm., January 2010). This allowed an inter-laboratory cross calibration of a kind commonly used in geochemistry because 14 additional Yunnan sites had already been collected by the Institute of Botany Chinese Academy of Sciences in Beijing and which overlapped the geographic range of the Kunming samples. The Kunming and Beijing sample sets plot in the same, but previously unoccupied, region of physiognomic space alongside the Beijing summer school samples and samples from India. This not only attests to the ability of different groups to collect and score independently, but collectively these samples represent a dataset that reflects physiognomic adaptations to the Asian monsoon.

2.3. Taphonomic Information Loss and the Need for Redundancy in Physiognomic Scoring: Assessing the contribution of leaf size to MAT and Enthalpy

The advantage of the multivariate approach becomes apparent when character data are missing, as they invariably are in fossils, because the positioning of an unknown in multidimensional space is determined by the array of surviving characters. The use of multiple characters inevitably carries

with it a degree of data redundancy, but this affords a robustness to the technique that is lacking in univariate methods.

Information loss can take the form of missing characters (e.g. leaf apices may be destroyed during pre-depositional transport processes, or after fossilization during the collecting process) or loss of species. Although some aspects of taphonomic information loss have been explored previously [Spicer et al., 2005] new analyses have been carried out and are reported here in respect of MAT and Enthalpy across the structure of PHYSG3BRC physiognomic space. Twenty six of the PHYSG3BR sites were selected to represent the spread across the multidimensional volume of physiognomic space, and from each of these 26 sites each character suite was removed sequentially. Thus in the first analysis all lobing information for all species in each site was removed. In the second analysis the lobing information was restored but the margin information was removed. This process continued across the scoring spectrum of all 7 character suites. Sites degraded in this way were treated as new passive unknowns and estimates for MAT and Enthalpy were derived using the standard PHYSG3BRC calibration and regressions. The results are given in Figures 1 and 2.

This process represents a “worst case” taphonomic loss for each character suite, not only because it was applied across all constituent taxa in any given sample, but because in some cases if all information from a character suite were missing the sample would be rendered unscorable. This is particularly so in the case of size information where, if it were all missing, so too would all margin, shape, apex, base etc. information. Nevertheless it provides a useful, if extreme and hypothetical, assessment of the contribution of each character suite in the context of a truly multivariate analysis where climate information still resides within surviving characters.

To test the effect of the loss of leaf size on MAT and enthalpy accuracy and precision, 32 sites were selected from the PHYSG3BRC calibration data set. These were also a subset of the 38 modern test

sites used by Peppe et al. [2010] obtained from the CLAMP website but excluded sites belonging to the "alpine nest" of Wolfe [1993] that are not appropriate for analysis using the PHYSG3BRC dataset because they lie outside physiognomic space as defined by that dataset. All leaf size data were then removed from each of the sites in the set of 32 and analyzed as passive "unknowns" using the standard PHYSG3BRC calibration and regressions. This exercise was similar to that of Peppe et al. [2010] but because the size information was treated as missing, instead of being scored differently from the calibration sites, no recalibration was required. The new predictions for MAT and enthalpy were then compared to the observed MAT and enthalpy at those sites, as well as the values predicted when all size data were present. The results are shown in Tables 2 and 3. In Table 2, although the mean value for MAT changes the uncertainty associated with the measurement does not. In Table 3 both the mean and the uncertainty change and those samples lacking size data demonstrate an overall rise in enthalpy uncertainties, as measured by the standard deviation of revised estimates, of just 1 kJ/kg.

2.4. Uncertainty in Paleoelevation Estimates of the Tibetan Plateau Associated with Leaf Size.

In their paper Peppe et al. [2010] cast doubt on the application of CLAMP to paleoaltimetry and cite as an example the work of Spicer et al. [2003] who applied the technique to a fossil locality in the southern Tibetan Plateau. Here a fossil site, well dated at 15 Ma by using single crystal $^{40}Ar/^{39}Ar$ analysis, yielded over 400 specimens assignable to 34 woody dicot leaf taxa with a completeness statistic [Herman and Spicer, 1997], *C*, of 0.849. *C* is given by:

$$C = F - M/P \quad (2)$$

where *F* is the actual number of data matrix cells filled when scoring a fossil assemblage, *M* is the minimum number of data matrix cells missing, and *P* is the minimum number of data matrix cells

that would be filled if all character states were to be scored for all taxa. *C* ranges from 0 to 1. The Namling value is well above the 0.66 recommended as the minimum required to yield a reliable analysis and the Namling site plots well within the boundaries of PHYSG3BRC physiognomic space, lying close to cool temperate sites from N.E. North America. The original analysis with the PHYSG3BR and MET3BR calibration datasets yielded a MAT of 8.1 ± 2.3 °C and an enthalpy estimate of 290 ± 6.4kJ/kg. Uncertainties are expressed as 2 standard deviations of the scatter of the residuals about the regression.

Because surface moist static energy is in part topography dependent, Spicer et al. [2003] used a climate model, constrained by oxygen isotope-derived SSTs and CLAMP enthalpy estimates from coeval sea level floras in Japan, to determine enthalpy at the tropopause. Model experiments showed enthalpy at the tropopause to be independent of the underlying surface height. Using the model enthalpy at the tropopause as a datum, enthalpy with respect to the fossil site was calculated. Peppe et al. [2010] question the age of one of the floras, but even if this were to be discarded there would be little effect on the model given the boundary conditions and the array of other data used to constrain it. Subsequent to the work of Spicer et al. [2003] pedogenic carbonates were recovered from the same section as provided the Namling flora [Currie et al., 2005]. Model results from δ^{18}Occ data, using the method of Rowley et al. [2001], indicate an elevation of 5200+1370/-605 m at 15 Ma, identical, within the quoted uncertainties, with the value determined from fossil leaf physiognomy. Such congruence supports the validity of elevation estimates derived from both methodologies, at least in this case. Furthermore several tectonic models that do not draw on those results require an early uplift of the south central plateau (e.g. England and Searle, 1986, England and Houseman, 1993, Murphy et al. 1997, Tapponier et al. 2001).

Despite this independent corroboration, Peppe et al. [2010] expressed concerns over the possible effect of “bias” in leaf size measurements pertaining to enthalpy, and thus paleoelevation

measurements. In the light of these concerns the Namling leaf data were re-analyzed three ways. Analysis 1) represents the original MET3BR climate calibration and the complete set of Namling leaf scores. This is the same analysis as reported in Spicer et al. [2003] using the PHYSG3BR dataset. Note, however, that this new analysis yields slightly different values from the original (e.g. 0.07°C difference in MAT and 0.2 kJ/Kg in Enthalpy) due to the correction of small errors found in the original PHYSG3BR data matrix [Spicer et al., 2009]. Analysis 2) is as for analysis 1), but with all nine size categories removed from the Namling scores. The PHYSG3BRC data were left intact so no regression recalibration was required. In Analysis 3) the original MET3BR climate calibration was used, but with all nine size categories removed from all taxa in the PHYSG3BRC calibration dataset as well as from the Namling scores. Inevitably this changed the calibration and new regressions were calculated. The results are given in Table 4, which shows that removal of size data from the Namling physiognomic scores and from the calibration data set leads to an increase in overall uncertainly of just 0.91 kJ/kg which when combined with the model uncertainties equates to an increased elevation uncertainty compared to when all the size data are present in both the calibration data and the fossil site of just ± 52 m at the 95 % confidence level. However the elevation estimate is reduced by 785 m.

3. Results and Discussion.

3.1 Uncertainties due to scoring

Concerns over the ability of different people to correctly interpret the scoring protocols led to the Xioalongmen experiment, the results of which are shown in Table 1. For MAT and enthalpy the standard deviation of the student results was less than that inherent in the CLAMP statistical methodology. This reflects the combined uncertainties inherent in both collecting and scoring. It seems then that the revised scoring documentation removes any significant user error, even in

situations where there might be a language divide. This is supported by the ability for different laboratories to independently successfully score numerous modern Chinese sites.

3.2. Assessing the role of leaf size in determining MAT and enthalpy estimates using modern test sites

A fundamental assumption underlying the approach of Peppe et al. [2010] is that the multiple interactions and correlation that connect foliar physiognomy and climate variables can be deconstructed to single leaf character/climate variable interactions. There is no theoretical or ecophysiological basis for this assumption. Moreover by ignoring published CLAMP protocols relating to minimum sample size their test data are rendered unreliable and will inevitably lead to elevated uncertainties. By changing scoring methodologies, and failing to recalibrate CLAMP when appropriate, their conclusions must further be called into question. Insufficient data have been made available for replication so a simpler way to test their findings is to examine the effect that complete removal of all size data are likely to have not only on enthalpy but also that other key climate variable used in paleoaltimetry - MAT.

Figures 1 & 2 show that for both the climate variables commonly used in paleoaltimetry, MAT and enthalpy, catastrophic loss of all leaf size information in a potential fossil assemblage leads to minimal loss of precision. A wider analysis shows that margin and apex information are far more important to the correct determination of enthalpy. Presumably this is because margin characters are strongly correlated with temperature and apex form is strongly correlated with precipitation and hence moisture. Temperature and moisture are the two critical elements defining moist static energy.

Leaf size is usually also correlated with moisture availability [Wilf et al. 1998], but in CLAMP leaf size is most important in assessing WMMT and 3-WET. Interestingly these, and relative humidity,

show the poorest physiognomy/climate relationships as measured by scatter about the regressions, and it may well be that different ways of recording size might improve this. However it is also worth remembering that when water is not limiting to plant growth leaf size is likely to be unconstrained, except by structural and resource investment issues, and will exhibit very poor correlation with any aspect of climate. Inevitably this introduces noise. Thus a *prima facie* case might be made for abandoning leaf size classification in CLAMP when the key objective is paleoaltimetry and this is tested in section 3.3.

In an attempt to partially replicate the Peppe et al. [2010] analyses of CLAMP sites that have significant elevation and enthalpy differences, 32 such sites from the PHYSG3BRC data set where analyzed as passive unknowns with all size data removed. Table 2 shows the resulting effects on MAT estimates and Table 3 shows the Enthalpy estimates. Further information on the locations and characteristics of these sites, including the score sheets, is available on the CLAMP website. It is evident from the right hand column of both tables that removal of size data has little effect on the ability of CLAMP to return reliable estimates as measured against the observed. When translated to elevation the mean difference between the enthalpy values obtained from samples lacking size data and that observed is just 49 m with a 2 sigma uncertainty of ± 63 m. The maximum difference is just 118 m.

3.3. Assessing the role of leaf size in determining MAT and enthalpy estimates using the Namling fossil site.

Despite the experiments in section 3.2 indicating negligible effect from the loss of size data it is worth examining the effect of removing leaf size information from an actual fossil leaf assemblage that has been used in palaeoaltimetry and for which there is independent corroborative evidence for paleoelevation. Such a site is that at Namling. In Table 4 it is evident that removal of the size information from all species in the Namling assemblage (a form of extreme taphonomic filtering)

yields an altitude estimate only 97 m lower from that with the character suite intact. Because the calibration did not require changing in this instance the uncertainties (a combination of model uncertainties and 2 sigma CLAMP uncertainties) remain the same at ± 634 m. However it is also evident from Figure 2 that the slope and intercept of the regression line is changed slightly for passive samples lacking size data. If the Namling site is "corrected" using the equation of this regression line the revised enthalpy estimate is 289.6 kJ/km and the elevation estimate increases to 4.68 km, or 60 m higher than when fossil leaf size data are included.

To test the effect of completely removing all size data from CLAMP size information was eliminated from both the PHYSG3BRC calibration data as well as from the Namling scores. Here the estimated Namling elevation drops by 785 m compared to the complete data sets. However the uncertainty rises only by an additional 52 m. Thus complete elimination of all size data still yields height estimates close to those obtained by those retaining size data and the upper elevation of the CLAMP estimate lacking all size data still overlaps with the lower estimate derived by Currie et al. [2005] using oxygen isotopes. Note that a regression "correction" directly comparable to that derived from Figure 2 is not possible here as all sites lacked size information. A different correction factor can be derived from the regression of the observed enthalpy of the calibration sites v the predicted value for those sites. In this case the slope is lower and the intercept higher than when the complete data are used and this yields a much reduced paleoelevation estimate of around 3.61 km. This suggests that either oxygen isotopes greatly exaggerate paleoelevation or that size data should be retained at least for CLAMP calibration purposes.

4. Conclusions

Tests with novice scorers and inter-laboratory independent scoring indicate that, with the amendments made to the scoring documentation, scorer error is less likely to be an issue than it once was and the best current estimates of uncertainties arising from this source is that it is less than

the uncertainties inherent in the 'noise" in the calibration data sets. Unfortunately in their critical appraisal of CLAMP uncertainties arising from size measurements Peppe et al. [2010] appear not to have followed CLAMP protocols and did not recalibrate CLAMP when appropriate.

All of the claims of Peppe et al. [2010] concerning alleged "bias" in scoring leaf size for CLAMP analyses, and consequential biases in enthalpy and height estimates, are unhelpful due to failure to recalibrate CLAMP when test samples were scored differently from samples used for CLAMP calibration. In some instances the resulting uncertainties were also inflated by using test samples that did not conform to CLAMP protocols regarding 1) capturing all morphological variation within a species, 2) sample species diversity or 3) collection, although this last point remains unclear due to lack of detail concerning collection procedure.

Although the Peppe et al. [2010] experiments cannot be repeated with the appropriate re-calibrations due to the data being unavailable, they have highlighted the role of leaf size in CLAMP paleoelevation analyses and simple tests excluding all size data are illuminating. The exclusion of all size data from modern or fossil passive test samples results in CLAMP estimates of MAT and enthalpy that are statistically identical (within uncertainty) to those obtained when size data are included. Estimates differ and uncertainties increase when all size data are excluded from both test samples and the calibration scheme, but even then in the case of the fossil Namling site the prediction of paleoelevation was associated with a minimal loss of precision.

Loss of calibration data is always likely to lead to higher uncertainties and for this reason removal of size data from CLAMP calibrations is not to be recommended. However, because leaf size data is inherently noisy its loss, coupled with the multivariate nature of the physiognomy/climate relationships, does not seriously compromise CLAMP's ability to deliver an estimate of enthalpy and thus surface height, derived from fossil floras that may not fully represent the range of leaf

sizes present in the once living source vegetation. Even in the extreme case of all size data being absent in a fossil flora the resulting uncertainties remain well within the range regarded by Forest et al. [1999], as being important. CLAMP would appear to be highly robust in respect of paleoaltimetry determinations even in the highly unlikely case of total leaf size data loss from a fossil assemblage Rather than dismiss a potentially useful paleoaltimeter, recalibrating CLAMP using measured leaf sizes could be done provided recalibration took place. However, given the relative immunity to leaf size in deriving MAT and enthalpy for CLAMP it is doubtful that the effort would be repaid by significant improvements in precision.

This controversy over foliar physiognomic methodology, and in particular the capturing of leaf size information, is not new. Nor is the attempt to deconstruct the complexity of leaf physiognomic/environmental relationships. In an exchange that followed an earlier attempt by Wilf et al. [1998] to criticize CLAMP methodology and deconstruct character/climate relationships using simple univariate linear regressions, Wolfe and Uemura [1999, p.91] noted “The complexity of the interactions of various environmental parameters that produces various compromises in physiognomic adaptations demands a more sophisticated approach than presented by Wilf et al.”. Wilf et al. [1999, p.92] responded: “ With regard to the suitability of a univariate approach, we acknowledge that a method that simultaneously considers the effects on leaf size and shape of many variables such as precipitation, temperature, seasonality, soil characteristics, etc. might be more desirable.”

CLAMP, with or without leaf size data, appears to be such a method but it is not, and should not be, a static methodology. Improvements are needed and are being made, for example in terms of increasing the geographic spread of the calibration data sets, but when testing CLAMP to achieve such improvements it is essential that valid test data are used and protocols followed.

Acknowledgments.

We thank staff and students of the Beijing Summer School and for the support provided by Research Councils UK and the Chinese Academy of Sciences. This work was conducted under the terms of a Visiting Professorship for Senior International Scientists awarded to RAS by the Chinese Academy of Sciences, and a Royal Society/Chinese Academy of Sciences Joint Project .

References

Ackerly, D.D., S.A. Dudley, S.E. Sultan, J. Schmitt, J.S. Coleman, C.R. Linder, D.R. Sandquist, M.A. Geber, A.S. Evans, T.E. Dawson, and M.J. Lechowicz (2000), The evolution of plant ecophysiological traits: recent advances and future directions, BioScience, 50, 979–995.

Axelrod, D.I. (1968), Tertiary floras and topographic history of the Snake River Basin, Idaho, Geol. Soc.Amer. Bull., 79, 713-734.

Bailey, I. W. and E. W. Sinnott (1915), A botanical index of Cretaceous and Tertiary climates, Science, 41, 831-834.

Bailey, I. W. and E. W. Sinnott (1916), The climatic distribution of certain types of angiosperm leaves, Am. J. Bot., 3, 24-39.

Benzecri, J.P. (1973), L'analyse des données: L'analyse des coreespondeneces. Dunod, Paris, 619 pp.

Currie, B. S., D.B. Rowley and N.J. Tabor (2005), Middle Miocene paleoaltimetry of southern Tibet: Implications for the role of Mantle Thickening and delamination in theHimalayan orogen, Geology, 33, 181-184.

Doyle, J.A. (2007), Systematic value and evolution of leaf architecture across the angiosperms in light of molecular and phylogenetic analysis, Cour. Forschung. Senck., 258, 21-37.

England, P.C., and P. Molnar (1990), Surface uplift, uplift of rocks, and exhumation of rocks. Geology, 18, 1173-1177.

England, P. C., and M. Searle (1986), The Cretaceous-Tertiary deformation of the Lhasa block and its implications for crustal thickening in Tibet. Tectonics, 5, 1-14.

Forest, C.E., J.A. Wolfe, P. Molnar, and K.A. Emanuel (1999), Palaeoaltimetry incorporating atmospheric physics and botanical estimates of palaeoclimate. Geol. Soc. Amer. Bull. 111(4), 497-511.

Garzione, C. N., D. L. Dettman, J. Quade, P. G. DeCelles, and R. F. Butler (2000), High times on the Tibetan Plateau: Paleoelevation of the Thakkhola graben, Nepal, Geology, 28, 339-342.

Gosh, P., C.N. Garzione and J.M. Eiler (2006), Rapid uplift of the Altiplano revealed through ^{13}C – ^{18}O bonds in Paleosol Carbonates. Science, 311, 511-515. DOI: 10.1126/science.1119365.

Green, W.A. (2006), Loosening the CLAMP: an exploratory graphical approach to the Climate Leaf Analysis Multivariate Program, Palaeo. Elect. 9. 9A: 17p.

Hellicker, B.R., and S.L. Richter (2008), Subtropical to boreal convergence of tree-leaf temperatures, Nature, 454, 511–514. DOI: 10.1038/**nature**07031.

Herman, A.B., and R.A. Spicer (1997), New quantitative palaeoclimate data for the Late Cretaceous Arctic: evidence for a warm polar ocean, Palaeogeog., Palaeoclimatol., Palaeoecol., 128, 227-251.

Hill, M.O. (1973), Reciprocal Averaging: an eigenvector method of ordination, J.Ecol., 61, 237-249.

Hill, M.O. (1974), Correspondence Analysis: a neglected multivariate method, Appl. Statist., 23, 340-354.

Hill, M.O. (1991), Patterns of species distributions in Britain elucidated by canonical correspondence analysis, J. Biog. 18, 247-255.

Hill, M.O. and H.G. Gauch (1980), Detrended correspondence analysis: An improved ordination technique, Vegetatio, 42, 47-58.

Jordan, G. J. (1997), Uncertainty in palaeoclimatic reconstructions based on leaf physiognomy. Aust. J. Bot., 45, 527-547.

Kennedy, E.M., R.A. Spicer, and P.M. Rees (2002), Quantitative palaeoclimate estimates from Late Cretaceous and Paleocene leaf floras in the northwest of the South Island, New Zealand. Palaeogeog., Palaeoclimatol., Palaeoecol., 184, 321-345.

Lande, R. and S.J. Arnold (1983), The measurement of selection on correlated characters, Evolution 36, 1210–1226.

Little, S.A., S. Kembel, P. Wilf and D. Royer (2008), Phylogenetic signal in leaf traits and its influence on leaf-climate correlations, Botany 2008: Annual Meeting of the Botanical Society of America. Abstract, 626.

McElwain, J.C. [2004], Climate-independent paleoaltimetry using stomatal density in fossil leaves as a proxy for CO_2 partial pressure, Geology, 32, 1017–1020.

Meyer, H. W. (1992), Lapse rates and other variables applied to estimating paleoaltitudes from fossil floras, Palaeogeogr. Palaeoclim. Palaeoecol., 99, 71-99.

Palmer, M. (1993), Putting things in even better order: the advantages of canonical correspondence analysis, Ecology, 74, 2215-2230.

Peppe, D.J., D.L. Royer, P. Wilf and E.A. Kowalski (2010), Quantification of Large Uncertainties in Fossil Leaf Paleoaltimetry, Tectonics, 29, TC3015, doi:10.1029/2009TC002549

Pollisar, P.J., K.H. Freeman, D.B. Rowley, F.A. Mcinerny, and B.S. Currie (2009), Palaeoaltimetry of the Tibetan Plateau from D/H ratios of lipid biomarkers, Earth Planet. Sci. Lett., 287, 64-76. DOI:10.1016/j.epsl.2009.07.037.

Povey, D. A. R., R. A. Spicer and P.C. England (1994). "Palaeobotanical investigation of early Tertiary palaeoelevations in northeastern Nevada: initial results." Rev. Palaeobot. Palynol., 81, 1-10.

Quade, J., C. Garzione, and J. Eiler (2007), Paleoelevation reconstruction using pedogenic carbonates, Reviews in Mineralogy and Geochemistry, 66, 53-87.

Sahagian, D.L., A.A. Proussevitch, and W.D. Carlson (2002), Analysis of vesicular basalts and lava emplacement processes for application as a paleobarometer/paleoaltimeter, J. Geol., 110, 671–685.

Rowley, D.B., and C.N. Garzione, (2007), Stable isotope-based paleoaltimetry, Annu. Rev. Earth Planet. Sci. 35. DOI:10.1146/annurev.earth.1135.031306.140155.

Rowley, D.B., R.T. Pierrehumbert, B.S. Currie (2001), A new approach to stable isotope-based paleoaltimetry: implications for paleoaltimetry and paleohypsometry of the High Himalaya since the Late Miocene, Earth Planet. Sci. Lett., 188, 253–268.

Rowley, D.B. and B.S. Currie (2006), Palaeo-altimetry of the late Eocene to Miocene Lunpola basin, central Tibet, Nature, 439, 677-681. DOI: 10.1038/**nature**04506.

Spicer, R.A. (2000), Leaf Physiognomy and Climate Change, in *Biotic Response to Global change: the Last 145 Million Years*, edited by S.J. Culver and P. Rawson, P. pp. 244-264, Cambridge University Press, Cambridge.

Spicer, R.A. (2007), Recent and Future Developments of CLAMP: Building on the Legacy of Jack A. Wolfe. Cour. Forsch. Senck., 258, 109-118.

Spicer, R.A. (2008), CLAMP. In: Gornitz, V. (Ed.), Encyclopedia of Paleoclimatology and Ancient Environments. Springer, Dordrecht, pp. 156–158.

Spicer, R.A., N.B.W. Harris, M. Widdowson, A.B. Herman, S. Guo, P.J. Valdes, J.A. Wolfe, and S.P. Kelley (2003), Constant elevation of Southern Tibet over the past 15 million years, Nature, 421, 622-624.

Spicer, R.A., A.B. Herman, and E.M. Kennedy (2005), The Sensitivity of CLAMP to Taphonomic loss of Foliar Physiognomic Characters, Palaios, 20, 429-438. DOI 10.2110/palo.2004.P04-63

Spicer, R.A., P.J. Valdes, T.E.V. Spicer, H.J. Craggs, G. Srivastava, R.C. Mehrotra, and Jian Yang (2009), New development is CLAMP: calibration using global gridded meteorological data. Palaeogeog., Palaeoclimatol., Palaeoecol., 283, 91- 98. DOI:10.1016/j.palaeo.2009.09.009

Tapponnier, P., Zhiqin Xu, F. Roger, B. Meyer, N. Arnaud, G. Wittlinger, and Jingsui Yang. (2001), Oblique stepwise rise and growth of the Tibet plateau, Science, 294, 1671–1677. DOI: 10.1126/science.105978

Ter Braak, C.J.F. (1986), Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis, Ecology, 67, 1167–1179.

Weimann, M.C., S.R. Manchester, D.L. Dilcher, L.F., Hinojosa, and E.A. Wheeler (1998), Estimation of temperature and precipitation from morphological characters of dicotyledonous leaves. Am. J. of Bot., 85, 1796-1802.

Wilf, P. (1997), When are leaves good thermometers? A new case for Leaf Margin Analysis. Palaeobiol., 23, 373–390.

Wilf, P., S.L. Wing, D.R. Greenwood, and C.L. Greenwood (1998), Using fossil leaves as paleoprecipitation indicators: an Eocene example, Geology, 26, 203-206.

Wilf, P., S.L. Wing, D.R. Greenwood, and C.L. Greenwood (1999), Using fossil leaves as palaeoprecipitation indicators: An Eocene example: Reply, Geology, 27, p.92.

Wolfe, J.A. (1992), An analysis of present day terrestrial lapse rates in the western conterminous United States and their significance to palaeoaltitudinal estimates. U.S. Geol.l Surv. Bull. 1964, 1-35.

Wolfe, J.A. (1993), A method of obtaining climatic parameters from leaf assemblages, U. S. Geol. Surv. Bull., 2040, 1–73.

Wolfe, J.A., H.E. Schorn, C.E. Forest, and P. Molnar (1997), Palaeobotanical evidence for high altitudes in Nevada during the Miocene, Science, 276, 1672-1675.

Wolfe, J.A., C.E. Forest and P. Molnar (1998), Paleobotanical evidence of Eocene and Oligocene palaeoaltitudes in midlatitudes in western North America, Geol. Soc. Amer. Bull., 110, 664-678.

Wolfe, J.A. and K. Uemura (1999), Using fossil leaves as palaeoprecipitation indicators: An Eocene example: Comment, Geology, 27, p.91.

	Stdev Res	Pair 1	Pair 2	Pair 3	Pair 4	Pair 5	Pair 6	Pair 7	Pair 8	Pair Mean	Pair Stdev
MAT	1.72	11.19	11.41	9.91	13.07	14.79	10.69	12.48	9.08	11.58	1.71
WMMT	1.80	20.29	19.98	20.75	23.15	25.68	19.71	22.00	18.42	21.25	2.15
CMMT	2.54	2.69	3.57	-0.15	3.53	4.01	2.31	3.58	0.31	2.48	1.48
LGS	0.85	6.67	6.56	6.06	7.69	8.72	6.46	7.26	5.67	6.89	0.91
GSP	31.80	116.75	45.92	62.08	162.01	250.36	111.80	112.65	83.00	118.07	60.28
MM GSP	3.67	20.00	10.78	14.41	25.09	32.78	17.39	20.62	15.52	19.57	6.43
3-WET	13.81	69.34	35.01	45.69	89.06	122.00	63.32	69.72	52.88	68.38	25.49
3-DRY	8.99	52.88	28.00	42.35	67.96	87.28	41.68	57.66	40.38	52.27	17.47
RH	8.17	81.26	80.42	78.94	79.50	74.50	77.23	82.20	79.05	79.14	2.26
SH	0.98	9.78	9.74	8.44	9.55	8.57	8.65	10.29	8.57	9.20	0.67
ENTH	0.35	31.06	31.17	30.60	31.13	30.95	30.72	31.32	30.56	30.94	0.26

Table 1

Sample	MAT (°C)	MAT (°C)	MAT (°C)	MAT absolute difference	MAT absolute difference
	Observed	Predicted	Predicted minus size	Obs. minus predicted	Obs. minus predicted minus size
Guanica, Puerto Rico	26.8	23.4	25.3	3.4	1.5
Cabo Rojo, Puerto Rico	26.8	23.7	25.2	3.1	1.6
Borinquen, Puerto Rico	25.5	22.7	24.8	2.8	0.7
Keka, Fiji	25.2	25.8	26.7	0.6	1.5
Guajatica, Puerto Rico	24.8	23.4	23.1	1.4	1.7
Cabo San Lucas, Baja California	24.4	24.3	24.1	0.1	0.3
Alamos, Sonora	23.5	20.6	20.8	2.9	2.8
Empalme, Sonora	23.3	23.2	22.8	0.1	0.5
Buena Vista, Puerto Rico	22.0	23.7	25.7	1.7	3.7
Nagakubo, Yakushima	19.2	19.8	20.3	0.6	1.1
Yakusugi 260 m, Yakushima	17.9	17.2	17.1	0.7	0.8
Toro Negro, Puerto Rico	17.9	20.1	21.0	2.2	3.1
Childs, Arizona	17.9	19.2	18.0	1.3	0.1

Yakusugi 420 m, Yakushima	17.0	17.3	17.4	0.3	0.4
Portal, Arizona	15.8	17.8	18.0	2.0	2.2
Yakusugi 800 m, Yakushima	14.9	15.5	15.3	0.6	0.4
Santa Cruz, California	13.9	13.5	13.6	0.4	0.3
Placerville, California	13.9	11.2	11.1	2.7	2.8
Yakusugi 1080 m, Yakushima	13.4	12.6	12.5	0.8	1.0
Half Moon Bay, California	12.6	12.4	12.4	0.2	0.2
Powers, Oregon	12.0	10.6	10.0	1.4	2.0
Yakusugi 1350, Yakushima	11.9	11.3	11.0	0.6	0.9
North Bend, Oregon	11.5	11.4	10.7	0.1	0.8
Bandon, Oregon	11.0	11.9	11.3	0.9	0.3
Nestucca River, Oregon	10.4	9.0	7.7	1.4	2.8
Hood River, Oregon	10.3	8.2	7.5	2.1	2.8
Cape Blanco, Oregon	10.2	10.9	10.2	0.7	0.0
Three Lynx, Oregon	10.0	8.4	7.7	1.6	2.3

Clearwater, Washington	9.8	8.4	7.2	1.4	2.6
Parkdale, Oregon	8.5	7.8	7.0	0.7	1.5
Chuzenji-ko, Honshu	6.6	7.7	6.0	1.1	0.6
Republic, Washington	6.1	5.8	5.2	0.3	0.9
Suganuma, Honshu	4.0	4.5	3.0	0.5	1.0
			Mean	1.2	1.4
			Stdev	1.0	1.0

Table 2

Sample	Enth. (kJ/Kg)	Enth. (kJ/Kg)	Enth. (kJ/Kg)	Enth. (kJ/Kg)	Enth. (kJ/Kg)
				difference	difference
	Observed	Predicted	Predicted-Size	Observed minus	Observed minus
				predicted	predicted
					minus size
Guanica, Puerto Rico	338.5	335.6	342.9	2.9	4.4
Cabo Rojo, Puerto Rico	340.6	332.6	341.1	8.0	0.5
Borinquen, Puerto Rico	337.3	332.1	341.4	5.2	4.1
Keka, Fiji	337.7	340.8	337.1	3.1	0.6
Guajatica, Puerto Rico	338.5	334.2	326.9	4.3	11.6
Cabo San Lucas, Baja California Sur	328.8	322.8	334.5	6.0	5.7
Alamos, Sonora	321.7	320.0	323.4	1.7	1.7
Empalme, Sonora	325.3	316.1	326.9	9.2	1.6
Buena Vista, Puerto Rico	334.4	332.4	339.3	2.0	4.9
Nagakubo, Yakushima	316.4	318.3	316.0	1.9	0.4
Yakusugi 260 m, Yakushima	313.9	309.5	304.8	4.4	9.1
Toro Negro, Puerto Rico	326.0	322.1	322.1	3.9	3.9
Childs, Arizona	306.9	309.1	315.8	2.2	8.9

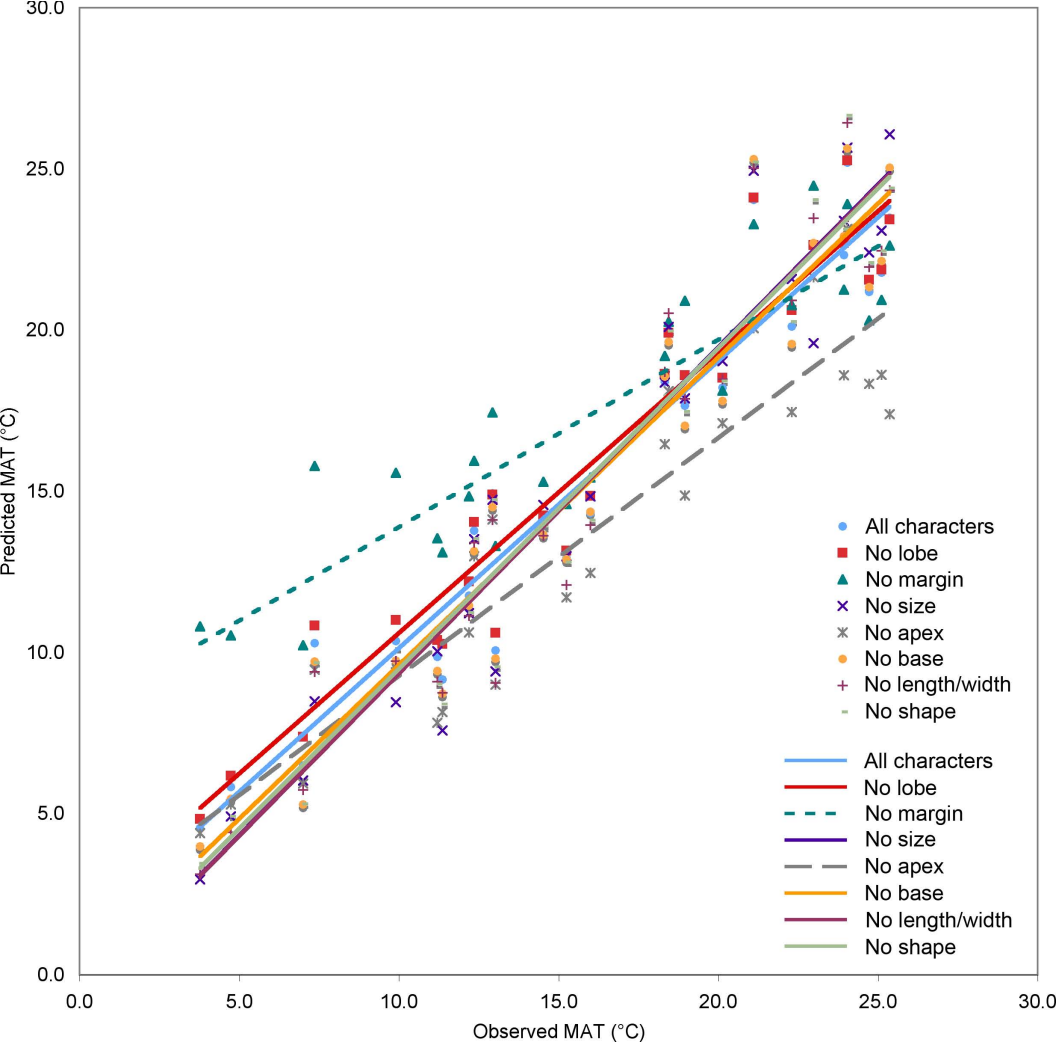
Yakusugi 420 m, Yakushima	312.4	308.7	304.7	3.7	7.7
Portal, Arizona	302.8	308.5	314.4	5.7	11.6
Yakusugi 800 m, Yakushima	308.6	307.6	304.4	1.0	4.2
Santa Cruz, California	307.6	306.2	306.8	1.4	0.8
Placerville, California	303.0	297.2	296.9	5.8	6.1
Yakusugi 1080 m, Yakushima	305.9	301.3	299.9	4.6	6.0
Half Moon Bay, California	307.4	303.6	304.7	3.8	2.7
Powers, Oregon	303.3	302.3	300.0	1.0	3.3
Yakusugi 1350, Yakushima	303.2	299.1	297.6	4.1	5.6
North Bend, Oregon	303.3	302.1	298.2	1.2	5.1
Bandon, Oregon	303.7	302.0	300.0	1.7	3.7
Nestucca River, Oregon	302.2	298.4	292.7	3.8	9.5
Hood River, Oregon	300.5	296.7	295.1	3.8	5.4
Cape Blanco, Oregon	303.4	300.8	298.6	2.6	4.8
Three Lynx, Oregon	299.1	296.3	294.8	2.8	4.3
Clearwater,	300.3	295.6	291.4	4.7	8.9

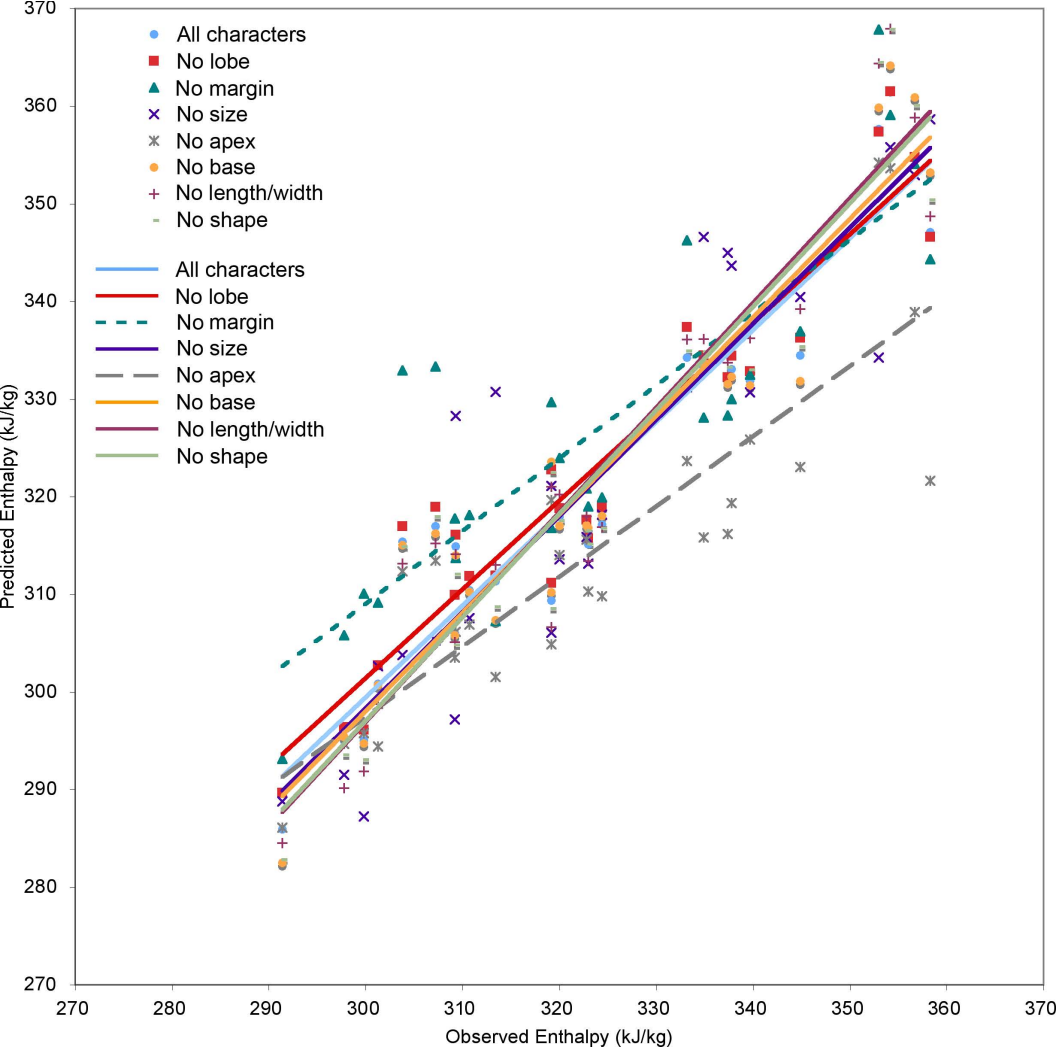
Washington					
Parkdale, Oregon	297.7	296.1	294.6	1.6	3.1
Chuzenji-ko, Honshu	293.7	300.9	293.4	7.2	0.3
Republic, Washington	293.4	289.6	289.0	3.8	4.4
Suganuma, Honshu	289.0	290.4	285.7	1.4	3.3
			Mean	3.7	4.8
			Stdev	2.1	3.1

Table 3

	Analysis 1	Analysis 2	Analyses 1 & 2	Analysis 3	Analysis 3
	Namling	Namling	STDEV	Namling	STDEV
MAT	8.080	7.821	1.181	8.662	1.204
WMMT	21.423	22.287	1.579	20.250	1.924
CMMT	-4.253	-5.473	1.898	-0.735	2.486
GROWSEAS	5.572	5.434	0.692	5.743	0.703
GSP	96.884	74.657	33.810	99.035	35.618
MMGSP	14.645	13.833	3.706	14.639	3.618
3-WET	54.177	47.244	14.112	54.094	13.713
3-DRY	35.780	38.999	9.358	32.312	8.936
RH	59.670	62.683	7.268	69.658	7.791
SH	3.326	3.792	0.920	5.798	1.192
ENTHAL	29.016	29.111	0.325	29.786	0.414
Elevation (km)	4.622	4.525		3.837	
Uncertainty (km)	± 0.634	± 0.634		± 0.686	

Table 4


30.0
25.0
20.0
15.0
10.0
5.0
0.0
Predicted MAT (°C)
0.0
5.0
10.0
15.0
20.0
25.0
30.0
Observed MAT (°C)
All characters
No lobe
No margin
No size
No apex
No base
No length/width
No shape
All characters
No lobe
No margin
No size
No apex
No base
No length/width
No shape


All characters
No lobe
No margin
No size
No apex
No base
No length/width
No shape
All characters
No lobe
No margin
No size
No apex
No base
No length/width
No shape
Predicted Enthalpy (kJ/kg)
Observed Enthalpy (kJ/kg)
270
280
290
300
310
320
330
340
350
360
370